The Role of Psychological Testing in Forensic Assessment

Kirk Heilbrun

Law and Human Behavior, Vol. 16, No. 3, Expert Evidence. (Jun., 1992), pp. 257-272.

Stable URL:
http://links.jstor.org/sici?sici=0147-7307%28199206%2916%3A3%3C257%3ATROPTI%3E2.0.CO%3B2-P

Law and Human Behavior is currently published by Springer.

Law and Human Behavior, Vol. 16, No. 3, 1992

The Role of Psychological Testing in Forensic Assessment

Kirk Heilbrun*

Despite the apparent widespread use of psychological tests in evaluations performed by psychologists to assist legal decision makers, there has been little critical but balanced examination of the appropriate parameters for the forensic use of such tests. The following discussion examines the nature of legal decision making, and concludes that the primary legal criterion for the admissibility of psychological testing is *relevance* to the immediate legal issue or to some underlying psychological construct. Assuming that *accuracy* is a more consistent concern for psychologists performing such evaluations, the criticisms of various commentators are discussed. Some criticisms appear appropriate and are incorporated into a set of proposed guidelines for the use of psychological tests in forensic contexts. Other criticisms appear misplaced, however, and the call for a whole sale ban on psychological testing in the forensic context is rejected.

The appropriate role of psychological testing in forensic assessment[1] has been debated for a number of years, and is far from clear at present. Critics have described such assessment procedures as "controversial" and "of doubtful validity and applicability in relation to forensic issues" (Ziskin, 1981a, p. 225; see also Faust & Ziskin, 1988, 1989; Ziskin & Faust, 1988). The major goal of this article is to discuss the research and commentary critical of psychological testing

* An earlier version of this paper was presented as part of a symposium on "Expert Testimony of Mental Health Professionals" (Randy K. Otto and Steve Berger, Co-Chairs) at the 1990 Annual Convention of the American Psychological Association, Boston, MA. Appreciation is expressed to Michael Perlin and Christopher Slobogin for their comments on an earlier draft of this paper and to Stephen Golding for his thoughts about the paper's organization and purpose. I would also like to thank Randy Otto for his role in organizing the APA symposium in which this paper was originally presented. Requests for reprints should be sent to Kirk Heilbrun, Department of Psychiatry, Medical College of Virginia, Box 710, Richmond, VA, 23298.

[1] For the purposes of this paper, *forensic assessment* will refer to information produced by mental health professionals intended for application to legal issues.

0147-7307/92/0600-0257$06.50/0

generally, and in the context of forensic assessment specifically, and to apply this discussion to the formulation of guidelines for the appropriate use of psychological testing in the forensic context.[2] Because this discussion will incorporate a number of ethical issues, I will include the most recent versions of existing ethical guidelines for forensic psychological assessment (Ethical Principles of Psychologists, 1991; Committee on Ethical Guidelines for Forensic Psychologists, 1991).

In the course of this discussion, psychological testing will be considered in its broadest sense. The focus will not be on particular instruments. Rather, the discussion will consider evidence relevant to a broad spectrum of tests, with the resulting guidelines applicable to psychological testing in this larger sense.

Much of the following discussion will necessarily focus on psychological assessment "in principle" rather than "in practice." I have been able to locate only one study providing empirical data on the normative uses of psychological tests in forensic assessment. In this study (Keilin & Bloom, 1986), a national survey was used to obtain information on the frequency with which 66 responding psychologists used psychological testing in child custody evaluations: some 75.6% reported using testing for the parents and 74.4% noted that they used psychological tests with the children. One of the ironies of the psychological testing controversy, however, is that ostensibly research-based criticism (e.g.. Faust & Ziskin, 1988) and responses (e.g., Fowler & Matarazzo, 1988) are made in the absence of virtually any normative data on the uses and abuses of psychological testing in the forensic context.

To do justice to this discussion, it will first be necessary to focus on the forensic assessment context itself. It differs in some important respects from the "therapeutic" contexts in which psychological testing is typically developed and employed. These differences have important implications for the purposes of mental health professionals' involvement, which in turn may affect whether (and what) psychological testing is used. Though there has been one previously published attempt to provide guidelines for the use of psychological tests in court (Blau, 1984), that effort relied heavily on the *Standards for Educational and Psychological Tests* in effect at that time (APA, 1974) and did not address in nature of legal decision making in any depth. The present article, by contrast, will first offer a discussion of legal decision making, and then provide parameters for the use of psychological tests in forensic assessment that are consistent with and flow directly from the nature of the legal decision-making process.

The Nature of Forensic Decision Making

What should legal decision makers seek from mental health professionals? Grisso's (1986) "model of legal competencies" contains the following elements: (a) functional abilities (abilities relevant for the legal competency in question), (b)

[2] To some extent, the *forensic context* discussed in this paper will focus more on criminal issues (e.g., trial competency, sanity at the time of the offense) than civil issues (e.g., personal injury litigation). Though the considerations discussed herein are intended to apply to both areas, the reader interested in a more detailed discussion of the application of psychological tests to child custody and personal injury litigation should consult Matarazzo (1990).

context (situation in which the competency must be demonstrated), (c) causal inference (nature of the relationship between the observed deficits and the legal ability), (d) interaction (between the person's particular abilities and the specific demands of the situation), (e) judgment (determination by the legal decision maker whether the person–situation incongruence is sufficient to warrant a finding of incompetency), and (f) disposition (the legal response to the individual authorized by the decision maker's finding).

This model describes a number of characteristics of forensic decision making. These characteristics will be used to structure various aspects of the following discussion. It should also be noted that forensic assessments are performed for legal decision makers and interpreted within applicable constitutional considerations and rules of evidence and procedure.

Psychological Test Solution: Relevance to the Legal Issue

The range of legal decisions that could potentially utilize mental health assessment is broad. Included are a host of civil,[3] criminal,[4] and family[5] issues (Melton, Petrila, Poythress, & Slobogin, 1987). For each of these legal issues, there is a standard derived from either statute or case law. Such standards vary in extensiveness and descriptiveness. For example, the standard for competency for execution ranges from a single word (e.g., insane, unfit) in 16 states[6] to the far more elaborate North Carolina standard:

> By reason of mental illness or defect he is unable to understand the nature and object of the proceedings against him, to comprehend his own situation in reference to the proceedings, or to assist in his defense in a rational or reasonable manner. (N.C. Gen.Stat. S 15A-1001, 1983).

Such differing legal standards can produce a different approach to selecting assessment tools, including psychological tests. Using the North Carolina standard, for example, the psychologist might select tests such as the WAIS-R to measure verbal intelligence, or the Boston Diagnostic Aphasia Examination, if aphasia were a possibility. These tests could be relevant to understanding and assisting counsel in a postcapital sentencing context, or to the psychological constructs underlying the functional abilities encompassed by the respective legal issues. When the standard is unelaborated, however (e.g., insane) the underlying psy-

---

[3] Civil issues would include competencies to consent to treatment and research, competency to enter into contractual relationships, civil commitment, guardianship, testamentary capacity, disability determination, and compensation for mental injuries.

[4] Criminal issues would include sanity at the time of the offense, diminished capacity, competencies to waive the rights to silence and to counsel, to stand trial, to be sentenced, to waive postconviction review, and to be executed.

[5] Family issues would include juvenile delinquency disposition, abuse and neglect, and child custody.

[6] States with a single-word standard for competency for execution ("insane") described in their respective state statutes include Alabama, Arizona, Arkansas, California, Connecticut, Maryland, Massachusetts, Mississippi, Nebraska, Nevada, New York, Rhode Island, and Wyoming. States with the identical standard taken directly from case law are Colorado, Kentucky, and Louisiana (Heilbrun, 1987).

chological or legal constructs are less clear. The absence of explicit guidance could result in selecting a broad range of psychological constructs to assess through testing, and such a decision would not necessarily be inappropriate.

Evidentiary and Constitutional Considerations: The Importance of Relevancy

Evidence offered by mental health professionals in legal proceedings is generally of the kind "reasonably relied upon" by professionals in the field (*Federal Rules of Evidence,* Rule 703). Some courts have also required that expert evidence meet the *Frye* test as being based on procedures that have gained "general acceptance" within a particular field (*Frye v. U.S.,* 1923), although many state courts now reject the *Frye* doctrine (M. L. Perlin, personal communication, July 20, 1990). The standard for admissibility of expert evidence has also been defined by using a relevancy analysis (Appelbaum, 1990; Slobogin, 1984). The issue of relevancy is the underlying predicate for all expert testimony, however, even when courts impose the additional demands that expert evidence be "reasonably relied on" or "generally accepted in the field" (Cleary, 1984; C. Slobogin, personal communication, August 16, 1990).[7]

Such relevancy could take two forms. An assessment instrument might directly measure the legal constructs underlying a given forensic issue. A number of such instruments have been developed in the past two decades, such as the Competency to Stand Trial Assessment Instrument (Laboratory of Community Psychiatry, 1973), the Interdisciplinary Fitness Interview (Golding, Roesch, & Schreiber, 1984), and the Rogers Criminal Responsibility Assessment Scales (Rogers, 1984a). These are discussed at length elsewhere (Grisso, 1986). A second kind of relevancy is seen when a psychological construct (such as intelligence) that is presumed to underlie part of a legal standard (such as "understanding of charges") is measured using a standard psychological test of intelligence, and the nature of that relationship is clarified in the report or through testimony. Particularly for criminal forensic issues, courts have not seemed inclined to limit the use of forensic instruments or psychological tests, so long as their relevance to the legal standard can be demonstrated. Such a demonstration could be made in the report itself or during direct testimony. This could be considered part of the ethical obligation incurred by a psychologist, working in a forensic context, to provide a "full explanation of the results of tests and the bases for conclusions . . . in language that the client can understand" (Committee on Ethical Guidelines for Forensic Psychologists, 1991, p. 664).

Flexible Standards in Legal Decision Making: Virtue, Not Error Variance

There is a flexibility within Anglo-American law, a capacity to consider political, moral, and community value influences, that is largely absent in the be-

[7] Whether "relevancy" and "general acceptance" are really interpreted in a different fashion by decision makers is an open question. It would appear that the relevancy standard is potentially broader; to the extent that this is true, the trend has been toward the admission of more, not less, expert evidence.

havioral sciences and their applied components such as clinical psychology. The diagnosis of a research subject in one cohort as schizophrenic and an identical subject in another cohort as affective disordered would be, for the purposes of behavioral science, considered error variance. In legal decision making, however, depending upon the larger political and moral context, such a discrepancy might be considered quite appropriate.

The argument has been made, for example, that the defense should be held to a less stringent standard of admissibility for mental health testimony on dangerousness than the prosecution, particularly in cases in which conviction could result in a death sentence (Slobogin, 1984). A more recent example of this position, again in the context of psychiatric predictions in death penalty cases, held that such predictions should be considered "an unconstitutional sword for the prosecution" but a "constitutional shield for the defense" (Wyda & Black, 1989). They buttress this argument with several precedents (*Chambers v. Mississippi,* 1973; *Rock v. Arkansas,* 1987), in which the Supreme Court upheld the right of defendants to present evidence (even "less reliable" evidence, such as hypnotically enhanced testimony) in the face of a state statute banning it. The Court reasoned that the defendant's Fourteenth Amendment right to present such evidence outweighed the authority of the state to bar it completely. It seems possible that the courts might adopt a similar approach to the admissibility of expert mental health testimony based on psychological testing, even if they were inclined to exclude some tests on the grounds of limited psychometric rigor. A different standard of rigor might be enforced depending, for example, on whether the evidence was being presented in a capital or a noncapital case, or by the defense or the prosecution.

This flexibility may work in the opposite direction as well. In civil litigation, for example, some courts have imposed a ban on the use of intelligence tests for placement of students in the school system. The use of such tests for vocational selection purposes has also been barred (Cohen, Montague, Nathanson, & Swerdlik, 1988; Elliott, 1987). In commenting, Meehl (1989) has noted that

> The law relies on hundreds of "generalizations" about human conduct . . . many of (which) would, if critically studied, turn out to be either false, or at least not highly generalizable from one situation to another. . . . By contrast, there are hundreds of research studies, in a variety of settings, involving many thousands of civilians and military personnel, in a variety of kinds of jobs, which show that proficiency at almost any kind of task will be correlated with the general intelligence factor. (pp. 544–545)

There is evidently a shifting standard with respect to the admissibility of behavioral science evidence. From the legal perspective, however, this can be accounted for by the nonscientific (i.e., political, moral, community standards) considerations so inherently a part of our system of law. One court, for instance, ruled that "when a program talks about labeling someone as a particular type and such a label could remain with him for the remainder of his life, the margin of error must be almost nil" (*Merriken v. Cressman,* 1973, p. 920). This would imply that a given test would need to have a validity coefficient approaching 1.00 to be selected, a criterion that no existing test can meet (Bersoff, 1982). What is clear is that some courts (e.g., *Griggs v. Duke Power Company,* 1971; *Larry P. v. Riles,*

1972, 1979) will apply this "shifting standard" to combat the invidious effects of de facto racial discrimination, even, for example, when test validity and reliability are high, if the impact of the resulting discrimination (or other problem) is judged to be sufficiently deleterious. It should thus be apparent that though a test's accuracy will be weighed by the court at the admissibility stage, the determination of accuracy may encompass the larger context of the litigation in addition to the test characteristics.

Outcome Measurement for Legal Determinations: Validation Problems

The prospect of using assessment instruments that have been validated against outcome data from forensic decision making is complicated by two characteristics of our legal system. The first involves the nature of the decisions themselves and the possibility of obtaining outcome data. Some issues embedded in legal standards do lend themselves to outcome research; one example is violent behavior. Although true experimental designs are rarely possible, given that judgments of dangerousness typically results in the detention of such individuals, it is feasible to perform research determining whether these individuals behaved violently in the institution.

For other legal standards, the determination of an outcome variable appropriate for research may be even more problematic (see Golding et al., 1984, for an extensive discussion of this problem). With an issue such as competency to stand trial, for example, the court's decision could be defined as the outcome measure. However, such decisions are heavily influenced by expert testimony (Melton et al., 1987). Alternatively, a "blue ribbon panel" composed of legal and mental health professionals familiar with competency to stand trial could provide the decision against which prediction could be calibrated (Roesch & Golding, 1980). Although the latter approach is methodologically preferable, it also increases the risk of expert testimony that invades the province of the decision maker (e.g., "the defendant is competent because my Inventory, validated against the judgments of a law professor, a practicing attorney, a psychologist, and a psychiatrist, says so"). Finally, it has also been proposed that incompetent defendants who wish to proceed with disposition of charges be allowed to do so despite their incompetency (Winick, 1987). Though such a procedure would remain problematic for a number of reasons, there is no doubt that it would dramatically improve the prospects for empirically calibrating legal decisions of trial competency against the genuine outcome of defendant behavior involved in the disposition of charges.

The second problem with calibrating legal decisions against base rate outcome data involves a policy debate regarding whether this kind of information should be used at all. Decision-theorists and social science researchers would advocate the use of such information (see, e.g., Arkes, 1989; Dawes, 1989; Faust, 1989; Haney, 1980; Meehl, 1989). However, some judges and legal scholars hold the opposite view, arguing that the use of mathematical probabilities in legal decision making should be prohibited because (a) the presumption of innocence

may be negated if the defendant is in a "high-risk" group, (b) the reliance on mathematical evidence diminishes the role of the jury, (c) it could increase the focus on more easily quantified variables, and (d) quantifying the probability of error in decision making seems intuitively unjust (Kaye, 1982; Tribe, 1971). It has also been argued that the use of empirically supported relationships between variables in the legal context would not be acceptable to courts without an underlying rationale for the relationship (Grisso, 1986).

Thus, both the nature of legal decision making and the debate about "mathematical" evidence may limit the use of well-validated instruments. Nonetheless, there is clearly a premium on correctness for the forensic decision maker. The decision will yield a disposition that is not subject to revision in the same fashion that clinical or scientific hypotheses may be (Grisso, 1986). As such, legal decision making by the factfinder should require the closest possible approximation to the truth, rather than hypotheses or speculations to be confirmed or refuted through further intervention.

Because of this premium on the accuracy of information provided to the factfinder, the results of psychological tests should not be used in isolation from history, medical findings, and observations of behavior made by others. This point has been made emphatically by Matarazzo (1990) in his discussion of forensic assessment of neuropsychological issues involved in personal injury and child custody litigation. It has also been made by others (e.g., Heilbrun, 1988; Melton et al., 1987; Rogers, 1986a; Shapiro, 1984, 1991) in reference to the assessment of criminal issues such as trial competency or mental state at the time of the offense. Impressions from psychological testing in the forensic context should most appropriately be treated as hypotheses subject to verification through history, medical tests, and third-party observations.

This "verification step" is crucial in forensic assessment for two reasons. First, psychological testing typically does not provide data that are directly relevant to the immediate legal issue. Rather, testing can provide information relevant to the threshold issue of mental or emotional disturbance; the causal connection between mental state and functional, legally relevant behavior remains to be assessed. Second, data obtained through psychological testing may, for a variety of reasons, provide an inaccurate representation of the individual. Such reasons include poor reliability or validity of the instruments, the deliberate attempt on the part of the individual to provide misleading information, nonstandardized test administration, and others. Verification of testing-based hypotheses through historical information, medical findings, and third party observations can significantly reduce such problems in relevance and accuracy.

The problem of outcome measures for legal determinations also suggests that legal constructs with potentially verifiable outcomes (such as dangerousness or treatability) should be investigated by social science researchers in order to provide outcome data, and mental health professionals should utilize such data in testimony. Even when outcome is less subject to verification, there is still a premium on accuracy in describing the psychological constructs underlying and relevant to the legal issue.

Psychological Testing: Guidelines for Use in Forensic Assessment

Though the uses of psychological testing in the context of legal decision making will inevitably be governed to some extent by legal policy and rules of evidence, mental health professionals need not be constrained from formulating their own standards for the uses of such testing in forensic settings. The preceding discussion has focused upon both legal and social science considerations relevant to the forensic use of psychological tests. In light of this discussion, I would propose the following guidelines to assist mental health professionals in determining whether a given psychological test should be used in a forensic evaluation.

Selection

(1) The test is commercially available and adequately documented in two sources. First, it is accompanied by a manual describing its development, psychometric properties, and procedure for administration. Second, it is listed and reviewed in Mental Measurements Yearbook *or some other readily available source.*[8]

There are a large number of categories of "psychological testing" that are potentially relevant to forensic decision making. A review of the Ninth Edition, Ninth Edition—Supplement, and Tenth Edition of the *Mental Measurements Yearbook* (Conoley, Kramer, & Mitchell, 1988; Conoley & Kramer, 1989; Mitchell, 1985), as well as the *Handbook of Scales for Research in Crime and Delinquency* (Brodsky & Smitherman, 1983) and *Evaluating Competencies* (Grisso, 1986) yielded 11 categories of psychological tests.[9] Given this broad range of areas, and the enormous number of tests (nearly 2,000) across all areas, it seems evident that this requirement in itself will not overly restrict the number of available instruments.

It is particularly important that these sources of documentation be available when psychological tests are used in the legal context. Psychologists should anticipate that data forming the bases for their conclusions should be available in order to allow opposing counsel adequate opportunity to prepare cross-examination and rebuttal (Committee on Ethical Guidelines for Forensic Psychologists, 1991). More generally, the availability of two published descriptions of a given psychological test will also permit opposing counsel to prepare arguments on the relevance and accuracy of this test. This is necessarily more specific, but still consistent with, the ethical guideline for documentation generally applicable

---

[8] Other sources might include reviews such as those by Brodsky and Smitherman (1983) or Grisso (1986). The purpose of this guideline is to advocate the use of instruments for which relevant information is readily available, not to increase the sales of the *Mental Measurements Yearbook.*

[9] The categories of psychological tests are as follows, with the number of tests in each area noted in parentheses: Achievement (75), Developmental (155), Educational (90), Intellectual (150), Miscellaneous (113), Neuropsychological (23, counting batteries as one each), Personality or "general scales" (578), Reading (144), Sensory–Motor (31), Speech/Hearing (90), and Vocations (510).

to psychologists. The most recent revision of the *Ethical Principles of Psychologists* merely notes that psychologists should "appropriately document their professional work . . . to ensure accountability, and to meet other requirements of institutions or the law" (1991, p. 32).

(2) Reliability should be considered. The use of tests with a reliability coefficient of less than .80 is not advisable. The use of less reliable tests would require an explicit justification by the psychologist.

The reliability of a psychological test should be considered before it is selected for use in a forensic assessment. A reliability coefficient in the .80's or .90's is usually considered desirable in psychological testing (Anastasi, 1988). There are a variety of ways in which to compute such a reliability coefficient, including test–retest, alternate-form, split-half, and Kuder–Richardson (for dichotomous items) or Cronbach's alpha (for multiple-scored items). The reliability figure just noted (.80) would refer primarily to test–retest reliability for measuring trait variables with objective tests, or interrater reliability for tests in which clinical judgment plays a significant part in data combination or interpretation. It could also refer to other indices of reliability, depending upon the test and the circumstances.

The larger relationship between reliability and validity can be described by using generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). The lower the reliability of a given test, the lower the limit on the validity of the construct being measured. It should thus be no surprise that tests with reliability coefficients below .80 have been criticized for containing excessive error variance and, hence, poorer validity. This is particularly true for many of the personality tests (Ziskin, 1981a). Reliability coefficients for psychological tests should be listed in their respective manuals, as well as in the *Mental Measurements Yearbook.*

(3) The test should be relevant to the legal issue, or to a psychological construct underlying the legal issue. Whenever possible, this relevance should be supported by the availability of validation research published in refereed journals.

Psychological tests may simply be irrelevant when clinicians attempt to use them in a straightforward fashion to measure legal concepts (Poythress, 1979; Ziskin, 1981a). The translation between the results of a test developed to measure a psychological construct such as "impulsivity" or "depression," and the outcome on a legal dimension such as "ability to assist counsel in one's own defense," *must* be justified by the *relevance* of that psychological construct to the legal construct.

Such justification should be made in the report, clarifying the evaluator's reasoning for selecting a given test on relevancy grounds. A justification can be made on theoretical grounds; if there is no research evidence with which to evaluate the accuracy or strength of the connection between psychological construct and legal issue, then the court should be so informed. However, in some instances, such evidence will be available. When a psychologist evaluates a defendant to assist the court in sentencing, for example, the construct of "psychopathy" would be useful; as measured by the Psychopathy Checklist (Hare, 1985), it has been related to success on parole (Hart, Kropp, & Hare, 1988; Serin, Peters

& Barbaree, 1990) and likelihood of criminal recidivism (Hare, McPherson, & Forth, 1988).

Administration

(4) Standard administration should be used, with testing conditions as close as possible to the quiet, distraction-free ideal.

An important consideration in the accuracy of psychological testing is standardization. Many tests have been developed with explicit, detailed instructions regarding administration. All tests share the need for a relatively quiet, distraction-free setting in administration. To the extent that the instructions and conditions of administration depart from the ideal in a given case, performance will be less than optimal. Also, tests have been normed under the conditions described in their manuals. When these conditions are not met in a single case, then the test norms may not apply in evaluating that individual's performance (Anastasi, 1988).

Interpretation

(5) Applicability to this population and for this purpose should guide both test selection and interpretation. The results of a test (distinct from behavior observed during testing) should not be applied toward a purpose for which the test was not developed (e.g., inferring psychopathology from the results of an intelligence test). Population and situation specificity should guide interpretation. The closer the "fit" between a given individual and the population and situation of those in the validation research, the more confidence can be expressed in the applicability of the results.

The standardization sample on which a given psychological test was developed may render it poorly generalizable to an individual in a forensic context (Tryon, 1979). In such instances, the evaluator may be confronted with an individual with characteristics from several populations (e.g., severely mentally ill, criminal offenders, neuropsychologically impaired) but not fitting well into any of them. With a test that has been well standardized and extensively researched (e.g., the MMPI), the evaluator may have the option of weighing the implications of the test results for several different populations. For most psychological tests, however, this luxury is not available. The evaluator must carefully scrutinize the standardization sample and subsequent research base for such tests to determine how close the "fit" between the present case and outcome samples, and qualify any testing-based conclusions accordingly.

Even when outcome data from similar populations are available, the behavioral sciences suffer from methodological problems that may make it difficult to generalize to the immediate forensic decision (Meehl, 1989). Additional problems with generalizability arise when an individual does not respond accurately to test questions (see Guideline below). Under some circumstances, the best data source may involve the use of specific information obtained and tabulated under clinical

conditions, so-called "quasi-experiments in real-life settings" (Meehl, 1989, p. 521).[10]

(6) Objective tests and actuarial data combination are preferable when there are appropriate outcome data and a "formula" exists.

Any discussion of validity in psychological testing must consider the clinical versus statistical prediction controversy. Originally described by Meehl (1954), the question was further refined by Sawyer (1966) into two components: (1) data collection (clinical versus objective), and (2) data combination (clinical versus actuarial). From the 18 studies available to Meehl in 1954, the number had grown to "about 100" by 1989, encompassing a range of tasks involving the diagnosis and prediction of human behavior; with a similar outcome obtained in virtually every study, "it is reasonable to conclude that the actuarial advantage is not exceptional but general and likely encompasses many of the unstudied judgment tasks" (Dawes, Faust, & Meehl, 1989, p. 1670). The use of actuarial data combination presupposes, however, that data have already been collected on similar patients, outcomes systematically measured, and predictor variables identified, optimally weighted, and cross validated. If no "formula" exists, then we have no alternative but to use our heads (Meehl, 1957).

(7) Response style should be explicitly assessed using approaches sensitive to distortion, and the results of psychological testing interpreted within the context of the individual's response style. When response style appears to be malingering, defensive, or irrelevant rather than honest/reliable, the results of psychological testing may need to be discounted or even ignored and other data sources emphasized to a greater degree.

Response style is another factor that may adversely affect the validity of conclusions that can be drawn from the results of psychological testing. Virtually every test requires that the patient attempt to accurately report his or her own experience. While mild forms of distortion or disengagement can be managed, the grosser forms of deception or failure to become engaged in the testing process can render test results meaningless. While the attempt to fabricate or greatly exaggerate psychopathology (malingering) has received most of the attention in the research and clinical literature, the assessment of response style should encompass all four response styles (1) *reliable/honest* (a sincere attempt is made to be accurate in responding, with factual inaccuracies attributable to poor understanding or misperception); (2) *malingering* (a conscious fabrication or gross exaggeration of physical and/or psychological symptoms, distinguished from factitious disorder in that the motivation for malingering goes beyond the desire to assume the patient role and is understandable in light of the individual's circumstances); (3) *defensive* (a conscious denial or gross minimization of physical and/or psychological symptoms, distinguished from ego defenses, which involve intrapsychic processes that distort perception); and (4) *irrelevant* (the failure to become engaged in the evaluation process; responses are not necessarily relevant to question content and may be random; Rogers, 1984b, 1988).

[10] For an example of this in practice, see Heilbrun (1990).

The only psychological test with extensive empirical support in measuring response style is the MMPI (Greene, 1988; Ziskin, 1981a, 1981b), although at least one specialized inventory, the Structured Inventory of Reported Symptoms (Rogers, 1986b), and clinical interviewing techniques (Resnick, 1987) have been developed to assess malingering as well. Relevant research on the falsified report of memory loss, with associated implications for interviewing, has also been described (Schacter, 1986). However, empirically validated techniques for detecting malingered neuropsychological deficits have apparently not been developed, despite reasonably convincing evidence that neuropsychologists cannot detect malingered symptoms of brain dysfunction using standard assessment techniques (Faust, Hart, & Guilmette, 1988; Heaton, Smith, Lehman, & Vogt, 1978; although cf. Bigler, 1990). Unless response style is explicitly measured and demonstrated to be honest, however, the interpretation of the results of psychological tests may be impossible. There is frequently an externally induced motivation to distort self-report in the forensic context; in one research report (Heilbrun, Bennett, White, & Kelly, 1990), the rate of reliable responding among mentally disordered offenders was only 33% of the total classified. Given this, the demand for assessing response style and validating self-report against third party information is high indeed (Heilbrun, 1988; Melton et al., 1987).

Critics of Psychological Testing in Forensic Assessment: Some Criticisms Are Misguided

Some additional comment on the appropriate conceptual approach to forensic assessment is in order, because critics who describe psychological testing as "of doubtful validity and applicability in relation to forensic issues" (Ziskin, 1981a, p. 225; see also Faust & Ziskin, 1988, 1989; Ziskin, 1969, 1975; Ziskin & Faust, 1988) seem to be noting the absence of a significant, direct relationship between psychological testing results and the immediate legal issue. This limited relationship has been observed by other investigators (Carbonell, Heilbrun, & Friedman, in press; Nicholson & Kugler, 1991) and should deter those who are sufficiently misguided to attempt to draw a direct connection between test results and the particular legal issue.

To draw such a direct connection, however, would involve a misunderstanding of the process of forensic assessment. Although such assessment is an applied rather than a scientific pursuit, it may be analogized to Meehl's description of "quasi-experiments in real-life settings" (1989, p. 521). Psychological testing can serve as one source of information that can both formulate and confirm or disconfirm hypotheses about psychological constructs relevant to the legal issue, but there are others as well: history, medical testing, interview data, and third-party observations of behavior can all be used for these purposes. Once hypotheses are formulated, they must then be translated into falsifiable terms, subject to verification, disconfirmation, or mixed support. It is noteworthy that such translation should not encompass the ultimate legal issue, which may have a host of nonscientific contributors such as political, social policy, moral, and community values (e.g., Morse, 1978, 1982; Slobogin, 1989) and thus be far less amenable to verification with mental health data.

Following the formulation of falsifiable hypotheses, the verification process can proceed much as it would in a scientific experiment. Does the defendant exhibit behavior consistent with the presence of the hypothesized psychological characteristic? (A researcher might call this *construct validity*.) Does the defendant show the absence of behaviors that are not consistent with the presence of the hypothesized construct? (We could analogize this to *discriminant validity*.) The remaining task is then to offer conclusions in terms that reflect the consistency of support for the hypothesis that was framed in psychological rather than legal terms (e.g., psychosis, cognitive awareness, and volition rather than insanity).

It is thus misguided to criticize psychological tests for being only weakly or indirectly related to legal issues. When used in the manner just described, it seems clear that the contribution of any test to a forensic assessment should be weighed according to how well it measures psychological constructs that are relevant to the legal issue and assists the evaluator in formulating and testing hypotheses.[11] While there are real problems with assessing certain legally relevant psychological constructs (e.g., volition), this is very different from saying that psychological tests do not relate very well to legal issues and thus should not be used to assist the courts in legal decision making. Properly used, psychological tests should not be expected to provide direct information about legal issues, and this criticism should only apply to evaluators who misuse them in such a fashion.

CONCLUSIONS

There are very limited empirical data on the uses of psychological testing in forensic contexts. This is clearly an area in need of research attention, to provide normative information on both the frequency with which various tests are used and the purposes for their use. The nature of legal decision making is such that the relevancy and, to a lesser extent, the accuracy of psychological tests will determine whether they will be admissible as a basis for expert testimony. These are reasonable criteria for mental health professionals as well as courts to use in the decision about the role of psychological testing in forensic assessment. It is clear, however, that courts will (and should) value and weigh accuracy differently than will psychologists.

It is therefore incumbent on psychology to develop an independent set of standards for the selection, administration, and interpretation of psychological testing in forensic contexts. Working as a group and a profession, psychologists

[11] This point was made succinctly by David Shapiro during the question period following the presentation of this paper and three others in a symposium at the 1990 American Psychological Association Convention in Boston. Addressing his question to the symposium discussant David Faust, who had emphasized the very limited value of psychological testing in providing valid information for forensic decision making, Dr. Shapiro asked, "Shouldn't we regard data from psychological tests—as from other sources we use in forensic assessment—as merely helping us form *hypotheses to be verified*?"

have recently developed a comprehensive set of ethical standards to guide the practice of forensic psychology (Committee on Ethical Guidelines for Forensic Psychologists, 1991). In a similar manner, it is hoped that the guidelines proposed in this article will focus attention and discussion on the parameters for using psychological testing in forensic assessment.

REFERENCES

American Psychological Association (1974). *Standards for educational and psychological tests*. Washington, DC: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Appelbaum, P. (1990, August). *Utility of psychiatric diagnosis in expert testimony*. Presented at the Annual Convention of the American Psychological Association, Boston, MA.

Arkes, H. R. (1989). Principles in judgment/decision making research pertinent to legal proceedings. *Behavioral Sciences & the Law, 7,* 429–456.

Bersoff, D. (1982). Regarding psychologists testily: The legal regulation of psychological assessment. In C. J. Scheirer & B. L. Hammonds (Eds.), *Psychology and the law* (pp. 37–88). Washington, DC: American Psychological Association.

Bigler, E. D. (1990). Neuropsychology and malingering: Comment on Faust, Hart, and Guilmette (1988). *Journal of Consulting and Clinical Psychology, 58,* 244–247.

Blau, T. H. (1984). Psychological tests in the courtroom. *Professional Psychology: Research and Practice, 15,* 176–186.

Brodsky, S. L., & Smitherman, H. O. (1983). *Handbook of scales for research in crime and delinquency*. New York: Plenum.

Carbonell, J., Heilbrun, K., & Friedman, F. (in press). Predicting who will regain trial competency: Initial promise unfulfilled. *Forensic Reports*.

Chambers v. Mississippi, 410 U.S. 284 (1973).

Cleary, E. W. (Ed.) (1984). *McCormick on evidence* (3rd ed.). St. Paul, MN: West.

Cohen, R. J., Montague, P., Nathanson, L. S., & Swerdlik, M. E. (1988). *Psychological testing: An introduction to tests and measurement*. Mountain View, CA: Mayfield.

Committee on Ethical Guidelines for Forensic Psychologists (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior, 15,* 655–665.

Conoley, J. C., & Kramer, J. J. (Eds.) (1989). *The tenth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.

Conoley, J. C., Kramer, J. J., & Mitchell, J. V. (Eds.) (1988). *The supplement to the ninth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Dawes, R. M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences & the Law, 7,* 457–467.

Dawes. R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Elliott, R. (1987). *Litigating intelligence: IQ tests, special education, and social science in the courtroom*. Dover, MA: Auburn House.

Ethical Principles of Psychologists (1991). *The APA Monitor, 22,* 30–35.

Faust, D. (1989). Data integration in legal evaluations: Can clinicians deliver on their premises? *Behavioral Sciences & the Law, 7,* 469–483.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241,* 31–35.

Faust, D., & Ziskin, J. (1989). Computer-assisted psychological evaluation as legal evidence: Some day my prints will come. *Computers in Human Behavior, 5,* 23–36.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake

believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56,* 578–582.

Fowler, R. D., & Matarazzo, J. D. (1988). Psychologists and psychiatrists as expert witnesses. *Science, 241,* 1143.

Frye v. United States, 293 F. 1013 (D.C. Cir 1923).

Golding, S. L., Roesch, R., & Schreiber, J. (1984). Assessment and conceptualization of competency to stand trial: Preliminary data on the Interdisciplinary Fitness Interview. *Law and Human Behavior, 9,* 321–334.

Greene, R. L. (1988). Assessment of malingering and defensiveness by objective personality inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 123–158). New York: Guilford.

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Grisso, T. (1986). *Evaluating competencies: Forensic assessments and instruments.* New York: Plenum.

Haney, C. (1980). Psychology and legal change: On the limits of a factual jurisprudence. *Law and Human Behavior, 4,* 147–199.

Hare, R. D. (1985). The Psychopathy Checklist. Unpublished manuscript (mimeo). Department of Psychology, University of British Columbia.

Hare, R. D., McPherson, L. M., & Forth, A. E. (1988). Male psychopaths and their criminal careers. *Journal of Consulting and Clinical Psychology, 56,* 710–714.

Hart, S. D., Kropp, P. R., & Hare, R. D. (1988). Psychopathy and conditional release from prison. *Journal of Consulting and Clinical Psychology, 56,* 227–232.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46,* 892–900.

Heilbrun, K. (1987). The assessment of competency for execution: An overview. *Behavioral Sciences & the Law, 5,* 383–396.

Heilbrun, K. (1988, March). *Third party information in forensic assessment: Much needed, sometimes collected, poorly guided.* Paper presented at the Mid-Winter Meeting of the American Psychology-Law Society/Division 41, Miami, FL.

Heilbrun, K. (1990). Response style, situation, third-party information, and competency to stand trial. *Law and Human Behavior, 14,* 193–196.

Heilbrun, K., Bennett, W. S., White, A. J., & Kelly, J. (1990). An MMPI-based empirical model of malingering and deception. *Behavioral Sciences & the Law, 8,* 45–53.

Kaye, D. (1982). The limits of the preponderance of the evidence standard: Justifiably naked statistical evidence and multiple causation. *American Bar Foundation Research Journal, 2,* 487–516.

Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice, 17,* 338–346.

Larry P. v. Riles, 343 F.Supp. 1306 (N.D. Cal. 1972) (order granting preliminary injunction) aff'd 502 F.2d 963 (9th Cir. 1974); 495 F.Supp. 926 (N.D. Cal. 1979) aff'd in part, rev'd in part *sub. nom. Larry P. by Lucille P. v. Riles.* 793 F.2d 969 (9th Cir. 1974).

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. American Psychologist, 45, 999–1017.

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268–273.

Meehl, P. E. (1989). Law and the fireside inductions (with postscript): Some reflections of a clinical psychologist. *Behavioral Sciences & the Law, 7,* 521–550.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1987). *Psychological evaluations for the courts: A handbook for mental health professionals and attorneys.* New York: Guilford.

Merriken v. Cressman, 364 F.Supp. 913 (E.D. Pa. 1973).

Mitchell, J. V. (1985). *The ninth mental measurements yearbook.* Lincoln, NE: University of Nebraska Press.

Nicholson, R. A., & Kugler, K. E. (1991). Competent and incompetent defendants: A quantitative review of comparative research. *Psychological Bulletin, 109,* 355–370.

Poythress, N. G. (1979). A proposal for training in forensic psychology. *American Psychologist, 34,* 612–621.

Resnick, P. (1987, October). *The detection of malingered mental illness.* Workshop presented at the American Academy of Psychiatry and Law, Ottawa, Canada.

Rock v. Arkansas, 107 S.Ct. 2704 (1987).

Roesch, R., & Golding, S. (1980). *Competency to stand trial.* Urbana, IL: University of Illinois Press.

Rogers, R. (1984a). *Rogers Criminal Responsibility Assessment Scales.* Odessa, FL: Psychological Assessment Resources.

Rogers, R. (1984b). Towards an empirical model of malingering and deception. *Behavioral Sciences & the Law, 2,* 93–111.

Rogers, R. (1986a). *Conducting insanity evaluations.* New York: Van Nostrand Reinhold.

Rogers, R. (1986b). *Structured Interview of Reported Symptoms (SIRS).* Clarke Institute of Psychiatry: Toronto, Unpublished scale.

Rogers, R. (1988) (Ed.). *Clinical assessment of malingering and deception.* New York: Guilford.

Sawyer, J. (1966). Measurement *and* prediction, clinical *and* statistical. *Psychological Bulletin, 66,* 178–200.

Schacter, D. L. (1986). Amnesia and crime: How much do we really know? *American Psychologist, 41,* 286–295.

Serin, R. C., Peters, R. D., & Barbaree, H. E. (1990). Predictors of psychopathy and release outcome in a criminal population. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 419–422.

Shapiro, D. L. (1984). *Psychological evaluation and expert testimony: A practical guide to forensic work.* New York: Van Nostrand Reinhold.

Shapiro, D. L. (1991). *Forensic psychological assessment: An integrative approach.* Boston: Allyn and Bacon.

Slobogin, C. (1984). Dangerousness and expertise. *University of Pennsylvania Law Review, 133,* 97–174.

Slobogin, C. (1989). The ultimate issue issue. *Behavioral Sciences and the Law, 7,* 259–268.

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84,* 1329–1393.

Tryon, W. W. (1979). The test-trait fallacy. *American Psychologist, 34,* 402–406.

Winick, B. (1987). Incompetency to stand trial: An assessment of costs and benefits and a proposal for reform. *Rutgers Law Review, 39,* 243–287.

Wyda, J., & Black, B. (1989). Psychiatric predictions and the death penalty: An unconstitutional sword for the prosecution but a constitutional shield for the defense. *Behavioral Sciences & the Law, 7,* 505–519.

Ziskin, J. (1981a). *Coping with psychiatric and psychological testimony* (Vols. 1–3, 3rd ed.). Marina Del Rey, CA: Law and Psychology Press.

Ziskin, J. (1981b). Use of the MMPI in forensic settings. *Clinical notes on the MMPI* (No 9). Minneapolis, MN: National Computer Systems.

Ziskin, J., & Faust, D. (1988). *Coping with psychiatric and psychological testimony* (Vols. 1–3, 4th ed.). Marina Del Rey, CA: Law and Psychology Press.